

Figure 1: Amino Acid and Nucleic Acid Sequences of *Hypoocrea jecorina* Cel7A

1	GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValLe.
101	. IAspAlaAsn TrpArgGTP ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys . CAGTCGGCT CCACTCTCCA ATCGGAGACT CACCGGCCCT CGAACAGCCTA ACTACGCTA CGAGCGAG CACGAACG CAGGACTGC TAGATGGCA GAAATGCTC TCTGGGGCA CCTGCACTCA ACAGACAGGC TCCGGGGCA
201	.AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrglyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla CGCGAAGAAC TGCTGCTGG ACCGTGGCG CTACGGCTCC ACGTACGGAG TTACACGAG CCGTAACAGC CTCTCCATTG GCTCTTGTCAC CCAGTCTGCG
301	GlnLysAsnVal GlyAlaArg LeuTyrlLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer . CAGAAGAACG TTGGCGCTCG CCTTACCTT ATGGCGAGCG ACACGACTTA CGAGGAATT ACCCTGCTTG GCAAGAGTT CTCTTTCGAT GTTGATGTTT
401	.SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr . CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTCTGT GTCCATGGAC GGGATGGTG GCGTGAACAA GTATCCCACC AACACCGCTG GCGCAAGTA
501	.GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn CGGCACGGGG TACTGTGACA GCCAGTGTTC CGCGATCTG AAGTTCATCA ATGCCAGGC CAACGTGAG GGCTGGGAGC CGTCACTCAA CAACGCCAAC
601	ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly . AGGGGATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCAACTC ATCTCCGAGG CTCTTACCCC CACCCCTTGC ACGACTGTCG
701	.GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrglyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyArg . GCCAGGAGAT CTGGGAGGGT GATGGGTGCG GCGGAACCTA CTCCGATAAC AGATATGGCG GCACTGGCA TCCCGATGGC TGCGACTGGC ACCATTAACCG
801	.LeuGlyAsn ThrSerPhePhe GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle CCTGGGCAAC ACCAGCTCT ACCAAGCTT ACCCTCGATA CACCAAGAA ATTGACCGTT GTCAAGCTTG AGGGCCCTG GGTGACGTC GGTGCGCATC
901	.GluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp . AACCGATACT ATGTCCAGAA TGGGGTCACT TTCCAGCAGC CCAACGCCGA GCTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACTGCG
1001	.AspAspTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu . AGGAGGCGAGA ATTGGGGGA TTCTCTTTCT CAGACAAGGG CGGGCTGACT CAGTCAAGA AGGCTACCTC TGGGGCATG GTGTTGGTC TGAGTCTGCG
1101	SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly . TCCGGGTGTC CCTGCTCAGT CGAATCTCAG TCTCCCAGG CCAAGGTAC CCTCTCCAAAC ATCAAAGTTCG GACCATTGG CAGCACGGGC AACCCCTGGCG
1201	.GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ProAlaThrThr ThrArgArg SerProGly ProThrGlnSer HistYrgly . GGGCAACCC TCCGGGGGA AACCCGGCTG GCACCAACC CAGCCCACTA CCTCTCCGGGA CCTACCCAGT CTCACACTACGG
1301	.GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrtTyrSer GlnCysIleu CCAGTGGGC CGTATTGGCT ACAGGGGCC CACGGTCTGC GCCAGGGCA CAACTTGCA GGTCCCTGAA CCTTACTACT CTCAGTGGCT G
1401	

Figure 2A: Multiple alignment of the CBH1 homologous sequences.

T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	1 (1) QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC (1) QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC (1) QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC (1) QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC (1) QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC (1) QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC	50
		51
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(51) YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA (51) YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA (51) YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA (51) YDGNTWSSTLCPDNECAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSQ (51) YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA (51) YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA	100
		101
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD (101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD (101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD (101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD (101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD (101) QKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD	150
		151
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(151) ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN (151) ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN (151) ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN (151) ADGGVSKYPSNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPESSNNAN (151) ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPESSNNAN (151) ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN	200
		201
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSND (201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYSND (201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYSND (201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQAICDGDGCGGTYSDD (201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDSCGGTYSGD (201) TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYS D	250
		251
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGAI (251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGAI (251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGAI (251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFATSGAI (251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGAI (251) RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGAI	300
		301
T. reesei mat H. orientalis mat H. schweinitzii mat T. koninlangbra mat T. pseudokoningii mat Consensus	(301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT (301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT (301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT (301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT (301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT (301) NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGGLT	350

Figure 2B: Multiple alignment of the CBH1 homologous sequences.

		351	
T. reesei mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H. orientalis mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H. schweinitzii mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
T. koninlangbra mat	(351)	QFKQATSGGMVLVMSLWDDYYANMLWLDSIYPTNETSSTPGAARGSCSTS	
T. pseudokoningii mat	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
Consensus	(351)	QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
		400	
T. reesei mat	(401)	SGVPAQVESQSPNAKVTFSNIKFCGPIGSTGNPSGGNPPGG-NPPGTTTR	
H. orientalis mat	(401)	SGVPAQLESQSPNAKVVYSNIKFCGPIGSTGNPSGGNPPGG-NPPGTTTR	
H. schweinitzii mat	(401)	SGVPAQLESQSPNAKVVYSNIKFCGPIGSTGNPSGGNPPGG-NPPGTTTR	
T. koninlangbra mat	(401)	SGVPAQLESQSTNAKVVFSNIKFCGPIGSTGNSSGGNPPGGGNPPGTTTR	
T. pseudokoningii mat	(401)	SGVPAQLESQSSNAKVVYSNIKFCGPIGSTGNSSGGSPGGGNPPGTTTR	
Consensus	(401)	SGVPAQLESQS NAKVVYSNIKFCGPIGSTGNPSGGNPPGG NPPGTTTR	
		450	
T. reesei mat	(450)	RPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. orientalis mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. schweinitzii mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTICASGTTCQQVLNEYYSQCL	
T. koninlangbra mat	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
T. pseudokoningii mat	(451)	RPATSTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
Consensus	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTQ VLNEYYSQCL	
		451	
T. reesei mat	(450)	RPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. orientalis mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. schweinitzii mat	(450)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTICASGTTCQQVLNEYYSQCL	
T. koninlangbra mat	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
T. pseudokoningii mat	(451)	RPATSTGSSPGPTQTHYGQCGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
Consensus	(451)	RPATTTGSSPGPTQTHYGQCGGIGYSGPTVCASGTTQ VLNEYYSQCL	
		499	

CGTCATCTCG	GCCTTCTTGG	CCACGGCCCG	TGCTCAGTCG	GCCTGCACTC	50
TCCAAACGGA	GACTCACCCG	TCTCTGACAT	GGCAGAAATG	CTCGTCTGGC	100
GGCACTTGCA	CCCAGCAGAC	AGGCTCCGTG	GTCATCGACG	CCAACTGGCG	150
CTGGACTCAC	GCGACTAAC	GCAGCACGAA	CTGCTACGAC	GGCAACACTT	200
GGAGCTCAAC	CCTATGCCCT	GACAACGAGA	CTTGCACGAA	GAATTGCTGC	250
CTGGACGGTG	CCGCCTATGC	GTCCACGTAC	GGAGTCACCA	CGAGTGCCGA	300
CAGCCTCTCC	ATCGGCTTCG	TCACGCAATC	TGCACAGAAG	AACGTTGGCG	350
CCCGTCTCTA	CCTGATGGCG	AGTACACGA	CTTACCAAGGA	GTTCACGCTG	400
CTTGGCAACG	AGTTCTCTT	TGACGTTGAT	GTTTCGCAGC	TGCCGTAAGT	450
GACAACCATT	CCCCGCGAGG	CCATCTCTC	ATTGGTTCCG	AGCTGACCCG	500
CCGATCTAAG	ATGTGGCTTG	AACGGCGCTC	TGTACTTCGT	GTCTATGGAT	550
GC GGATGGTG	GC GTGAGCAA	GTATCCCACC	AAACCCGCG	GC GCCAAGTA	600
CGGCACGGGC	TACTGCGACA	GCCAGTGCCC	CCGCGATCTC	AAGTTCATCA	650
ACGGCCAGGC	CAACGTTGAA	GGCTGGGAGC	CGTCCTCCAA	CAACGCCAAC	700
ACGGGTATTG	CGGGACACGG	AAGCTGCTGC	TCTGAGATGG	ATATCTGGGA	750
GGCCAACCTCC	ATCTCCGAGG	CTCTGACTCC	TCACCCCTGC	ACGACTGTTG	800
GCCAGGAGAT	CTGCGACGGT	GACGGCTGCG	GC GGAAACCTA	CTCCAACGAC	850
CGATATGGTG	GTACTTGC	GA TCCTGATGGT	TGTGATTGGA	ATCCATACCG	900
CTTGGGCAAC	ACCAGCTCT	ATGGCCCTGG	CTCGAGCTTC	ACCCTCGATA	950
CCACCAAGAA	GTTGACCGTT	GTCACCCAGT	TCGAGACCTC	GGGTGCCATC	1000
AACCGTTACT	ATGTCCAGAA	CGGCGTCACT	TACCAAGAAC	CCAACGCCGA	1050
GCTCGGTAGT	TACTCTGGTA	ATGAGCTCAA	CGATGACTAC	TGCACAGCTG	1100
AGGAGTCGGA	ATT CGGCGGC	TCCTCCTCT	CGGACAAGGG	CGGCCTTACT	1150
CAGTTCAAGA	AGGCCACTTC	CGGCGGCATG	GTCCTGGTCA	TGAGCTTGTG	1200
GGATGACGTG	AGTTGATAGA	CAGCATTAC	ATTGTCGTTG	GAAAGACGGG	1250
CGGCTAACCG	AGACATATGA	TATCTAACAG	TACTACGCCA	ACATGCTGTG	1300
GCTGGACTCC	ACCTACCCGA	CAAACGAGAC	CTCCCTCCACC	CCC GGCGCCG	1350
TGCGCGGAAG	CTGCTCCACC	AGCTCCGGCG	TCCCCGCTCA	GCTCGAGTCC	1400
CAGTCCCCCA	ACGCCAAGGT	CGTCTACTCC	AAACATCAAGT	TCGGGCCCAT	1450
TGGCAGCACC	GGCAACCCCA	GGCGCGGAAA	CCCTCCTGGC	GGAAACCCCTC	1500
CCGGCACCCAC	CACCACCCGC	CGCCCAGCTA	CCACCACTGG	AAGCTCTCCC	1550
GGACCTACTC	AGACTCACTA	CGGCCAGTGC	GGCGGCATCG	GCTACAGCGG	1600
CCCTACGGTC	TGCGCCAGCG	GCACGACCTG	CCAGG		1635

Figure 3: *H. oreintalis* genomic DNA sequence.

Figure 4A: *H. orientalis* amino acid signal sequence.

QSACTLQQTET HPSLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC	50
YDGNTWSSTL CPDNETCAKN CCLDGAAAYAS TYGVVTSADS LSIGFVTQSA	100
QKNVGARLYL MASDTTYQEF TLLGNEFSFD DVSQLPCGL NGALYFVSMD	150
ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFINGQANVE GWEPSNNAN	200
TGIGGGHGSCL SEMDIWEANS ISEALTPHPC TTVGQEICDG DGCGGTYSND	250
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKLLTV VTQFETSGAI	300
NRYYVQNGVT YQQPNAELGS YSGNELNDDY CTAEESEFGG SSFSDKGGLT	350
QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS	400
SGVPAQLESQ SPNAKVVYSN IKFGPIGSTG NPSGGNPPGG NPPGTTTTRR	450
PATTTGSSPG PTQTHYQCG GIGYSGPTVC ASGTTTCQVLN PYYSQCL	497

Figure 4B: *H. orientalis* mature amino acid sequence.

TCGGCCTGCA	CTCTCCAAAC	GGAGACTCAC	CCGTCTCTGA	CATGGCAGAA	50
ATGCTCGTCT	GGCGGCACTT	GCACCCAGCA	GACAGGCTCC	GTGGTCATCG	100
ACGCCAACTG	GCGCTGGACT	CACGCTACTA	ACAGCAGCAC	GAACTGCTAC	150
GACGGCAACA	CTTGGAGCTC	AACCCCTGTGC	CCTGACAATG	AGACTTGCGC	200
GAAGAACTGC	TGCCTGGACG	GTGCCGCTA	TGCGTCCACG	TACGGAGTC	250
CCACCGAGTGC	CGACAGCCTC	TCCATCGGCT	TCGTGACACA	GTCTGCACAG	300
AAAAACGTTG	GCGCCCGTCT	CTACCTGATG	GCGAGTGACA	CGACTTACCA	350
GGAGTTACG	CTGCTTGGCA	ACGAGTTCTC	ATTGACGTT	GATGTTTCGC	400
AGCTGCCGTA	<u>AGTGACAACC</u>	<u>ATTCCCCCGA</u>	<u>CGCCATCTTC</u>	<u>TCATTGGTTC</u>	450
<u>GAAGCTGACC</u>	<u>CGCCGATCTA</u>	<u>AGATGTGGCT</u>	<u>TGAACGGCGC</u>	<u>TCTTTACTTC</u>	500
GTGTCCATGG	ACGCAGATGG	TGGCGTGAGC	AAGTATCCCA	CCAACACCGC	550
CGCGGCCAAG	TACGGCACGG	GCTACTGTGA	CAGCCAGTGC	CCCCGCGATC	600
TCAAGTTTAT	CAACGGCCAG	GCCAACGTTG	AAGGCTGGGA	GCCGTCCTCC	650
AACAAACGCCA	ACACGGGTAT	TGGCGGACAC	GGAAGCTGCT	GCTCCGAGAT	700
GGATATCTGG	GAGGCCAACT	CCATCTCCGA	GGCTCTTA	CCTCACCCCT	750
GCACGAATGT	TGGCCAGGAG	ATCTGCGACG	GTGACGGCTG	CGGCGGAACC	800
TACTCCAACG	ACCGATATGG	TGGTACTTGC	GATCCTGATG	GTTGTGATTG	850
GAATCCATAC	CGCTTGGGCA	ACACCAGCTT	CTATGGCCCT	GGCTCGAGCT	900
TCACCCCTCGA	TACCACCAAG	AAAGTTGACCG	TCGTCACCCCA	GTTCGAGACT	950
TCGGGTGCCA	TCAACCGTTA	CTATGTCCAG	AATGGCGTCA	CTTACCAGCA	1000
ACCCAACGCC	GAGCTCGGCA	GTTACTCTGG	TAATGAGCTC	AACGATGCCT	1050
ACTGCACAGC	TGAAGAGTCG	GAATTGGCG	GTTCCCTCCTT	CTCGGACAAG	1100
GGCGGCCCTTA	CTCAGTTCAA	GAAGGCCACT	TCCGGCGGCA	TGGTCCCTGGT	1150
CATGAGCTTG	TGGGATGACG	<u>TGAGTCCATA</u>	<u>GAACAGCATT</u>	<u>CACATTGTCG</u>	1200
<u>TCGGAAAGAC</u>	<u>GGCGGCTAA</u>	<u>CCGAGACATT</u>	<u>ACAGTACTAC</u>	<u>GCCAACATGC</u>	1250
TGTGGCTGGA	CTCCACCTAC	CCGACAAACG	AGACCTCCTC	CACCCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAAGCTCC	GGCGTCCCAG	CTCAGCTCGA	1350
GTCCCAGTCC	GCCAACGCCA	AGGTCGTCTA	CTCCAACATC	AAGTTCGGAC	1400
CCATTGGCAG	CACCGGCAAC	CCCAGCGGG	GAAACCCCTCC	TGGCGGAAAC	1450
CCTCCCGGCA	CCACCACCCAC	CCGCGGCCCA	GCTACCACCA	CTGGAAGCTC	1500
TCCCGGACCT	ACTCAGACTC	ACTATGGCCA	GTGCGGCGGC	ATCGGCTACA	1550
GCGCCCTAC	GATCTGCGCC	AGCGGCACCGA	CCTGCCAGG		1589

Figure 5: *H. scweinitzii* genomic DNA sequence.

Figure 6A: *H. Schweinitzii* signal peptide.

QSACTLQETET HPSLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC	50
YDGNTWSSTL CPDNETCAKN CCLDGAAAYAS TYGVVTSADS LSIGFVTQSA	100
QKNVGarLYL MASDTTYQEF TLLGNEFSFD VDVSQQLPCGL NGALYFVSM	150
ADGGVSKYPT NTAGAKYGTG YCDSQCPRLD KFINGQANVE GWEPSNNAN	200
TGIGGHGSCC SEMDIWEANS ISEALTPHPC TNVGQEICDG DGCGGTYSND	250
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI	300
NRYYVQNGVT YQQPNAELGS YSGNELNDAY CTAEESEFGG SSFSDKGGLT	350
QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS	400
SGVPAQLESQ SANAKVVVYSN IKFGPIGSTG NPSGGNPPGG NPPGTTTTRR	450
PATTTGSSPG PTQTHYGQCG GIGYSGPTIC ASGTTCQVLN PYYSQCL	497

Figure 6B: *H. Schweinitzii* mature amino acid sequence. 497 residues

TCGGCCTGCA	CCATTCAAGC	GGAGACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCC	GGTGGTAGTT	GCACCTCGCA	AACCGGTTCT	GTGGTATTG	100
ACCGGAACTG	GCGATGGACT	CACCGCGACTA	ACAGCACCAC	GAACTGCTAC	150
GACGGTAACA	CTTGGGAGCTC	CAGTCTTGC	CCCGACAATG	AGAGTTGCGC	200
AAAGAACTGC	TGCCTGGACG	GTGCAGCCTA	CGCATCCACG	TACGGAGTCA	250
CCACGAGTGC	TGATAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTCAGCAG	300
AAGAAATGTTG	GCGCTCGTCT	CTACCTGATG	GCAAGCGACA	CGACCTACCA	350
GGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	TTTCGATGTT	GATGTTTCAC	400
AGCTGCCGT	AGTGA TAGC	ATTTACCTCC	GACGCCATCT	CATTGATTCC	450
CAGCTGACGG	CCAATTCAAG	ATGTGGCTTG	AACGGAGCCC	TTTACTTCGT	500
GTCCATGGAC	GCGGATGGTG	GCGTGAGCAA	GTATCCCTCC	AAACACTGCCG	550
GCGCCAAGTA	CGGCACGGGC	TACTGCGATA	GCCAGTGTCC	CCGTGATTG	600
AAGTTCATCA	ACGGCGAGGC	CAACGTTGAG	GGCTGGGAGC	CGGCTTCGAA	650
CAACGCCAAC	ACGGGTATTG	GCGGACACGG	AAGCTGCTGC	TCTGAGATGG	700
ATATCTGGGA	GGCCA ACTCC	ATCTCTGAGG	CCCTTACTCC	TCACCCCTGC	750
ACGACTGTG	GCCAGGCCAT	TTGCGATGGT	GACGGCTGCG	GTGGAACCTA	800
CTCCGATGAC	CGATATGGTG	GTACTTGC	TCCTGATGGC	TGTGACTGGA	850
ACCCATACCG	CTTGGGCAAC	ACCAGCTCT	ACGGCCCCGG	CTCGAGCTTC	900
ACCCCTCGACA	CCACCAAGAA	GATGACCGTC	GTCACCCAGT	TCGCTACTTC	950
GGGTGCCATC	AACCGATACT	ATGTCCAGAA	TGGCGTCACT	TTCCAGCAGC	1000
CCAACGCCGA	GCTCGGTAGC	TACTCTGGCA	ACACGCTCAA	CGATGCTTAC	1050
TGCGCAGCTG	AAGAGGCGGA	ATTCGGCGGA	TCATCTTCT	CAGACAAGGG	1100
TGGCCTTACC	CAATTCAAGC	AGGCTACTTC	AGGCGGCATG	GTCTTGGTTA	1150
TGAGCCTGTG	GGATGACGTG	AGTTCA GTA	TAGCATTGAC	ATTGTCGAGA	1200
GAACCATAGC	CGCTGACCGA	GACACAACAG	TACTACGCCA	ACATGCTGTG	1250
GCTGGACTCC	ATCTACCCGA	CGAACGAGAC	CTCCTCTACC	CCCGGTGCCG	1300
CGCGCGGAAG	CTGCTCTACC	AGCTCCGGTG	TCCCTGCCA	GCTCGAGTCT	1350
CAGTCTACCA	ACGCCAAGGT	CGTCTTCTCC	AACATCAAGT	TCGGACCCAT	1400
TGGCAGCACT	GGTAACTCCA	GCGGCGGAAA	CCCCCCGGGC	GGAGGAAACC	1450
CCCCCGGCAC	CACCACCA	CGACGCCAG	CTACCA	CGGAAGCTCT	1500
CCCGGACCTA	CTCAGACACA	CTATGGCCAG	TGCGGTGGAA	TTGGGTACTC	1550
GGGCCCCACG	GTCTCGGCCA	GCGGCAGCAC	ATGCCAGG		1588

Figure 7: *T. konilangbra* genomic DNA.

Figure 8A: *T. konilangbra* signal sequence.

QSACTIQAET HPPLTWQKCS SGGSCTSQTG SVVIDANWRW THATNSTTNC	50
YDGNTWSSL CPDNESCAKN CCLDGAAAYAS TYGVVTSADS LSIGFVTQSQ	100
QKNVGarLYL MASDTTYQEF TLLGNEFSFD DVSQLPCGL NGALYFVSMD	150
ADGGVSKYPS NTAGAKYGTG YCDSQCPRDL KFINGEANVE GWEPAASNAN	200
TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQAICDG DCGGGTYSD	250
RYGGTCDPDG CDWNPYRLGN TSXYGPGSSF TLDTKKMTV VTQFATSGAI	300
NRYYVQNGVT FQQPNAELGS YSGNTLNADY CAAEAEFGG SSFSDKGGLT	350
QFKQATSGGM VLVMMSLWDDY YANMLWLDSI YPTNETSSTP GAARGSCSTS	400
SGVPAQLESQ STNAKVVFSN IKFGPIGSTG NSSGGNPPGG GNPPGTTTR	450
RPATTTGSSP GPTQTHYGQC GGIGYSGPTV CASGSTCQVL NPYYSQCL	498

Figure 8B: *T. konilangbra* mature amino acid sequence.

CBH1 Homologs And Variant CBH1 Cellulases

Goedgebuur et al.

SN# Unassigned

Docket No. GC793-3

Sheet 10 of 14

TCGGCCTGCA	CCCTCCAGAC	GGAAACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCT	GGTGGCACTT	GCACCCAAACA	GACGGGCTCC	GTGGTCATCG	100
ACCGGAACGTG	GCGCTGGACT	CACGCTACGA	ACAGCAGCAC	GAACTGCTAC	150
GACGGTAACA	CTTGGAGCTC	AACCTTGTGC	CCTGACAATG	AGACTTGCAC	200
GAAGAACTGC	TGCTTGGATG	GTGCCGCCTA	CGCGTCGACG	TACGGAGTCA	250
CCACGAGCGC	TGACAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTGCCAG	300
AAGAATGTCG	GCGCCCGTCT	CTACTTGATG	GCGAGTGACA	CGACCTACCA	350
AGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	CTTCGATGTT	GATGTTTCCC	400
AGCTGCCGTA	<u>AGTGGCCAAC</u>	TACACCCCTT	<u>GACGGTATCC</u>	<u>TCTCATTGGT</u>	450
<u>TCCCAGCTGA</u>	<u>CTCGCGAAAT</u>	<u>TAAGATGTGG</u>	CTTGAACCGA	GCTCTTACT	500
TCGTGTCCAT	GGACGCGGAT	GGTGGCGTGA	GCAAGTATCC	CACAAACACT	550
GCCGGCGCCA	AGTACGGCAC	GGGTTACTGT	GACAGCCAGT	GCCCTCGTGA	600
TCTCAAGTTC	ATCAACGGCG	AGGCCAACGT	TGAGGGCTGG	GAGCCGTTCT	650
CCAACAACGC	CAACACGGGC	ATTGGCGGAC	ATGGAAGCTG	CTGCTCTGAG	700
ATGGATATCT	GGGAGGCCAA	CTCCATCTCT	GAGGCTCTTA	CTCCTCATCC	750
TTGCACGACC	GTCGGGCAGG	AAATTGCGA	TGGTGAATCC	TGCGGGCGAA	800
CCTACTCCGG	TGATCGATAT	GCGGGTACTT	GCGATCCTGA	TGGCTCGCAT	850
TGGAACCCAT	ACCGCTTGGG	CAACACCCAGC	TTCTACGGGC	CCGGCTCAAG	900
CTTCGCTCTT	GATACCACCA	AGAAGTTGAC	CGTTGTCACC	CAGTTGAGA	950
CTTCGGCGC	TATCAACCGG	TACTACGTCC	AGAATGGCGT	CACTTTCCAG	1000
CAGCCCAACG	CCGAGCTCGG	TAGTTACTCT	GGCAACTCGC	TCGACGATGA	1050
CTACTGCGCG	GCTGAAGAGG	CGGAGTTGG	TGGCTCTTCT	TTCTCGGACA	1100
AGGGCGGCCT	TACTCAATT	AAAAGGCTA	CTTCGGTGG	CATGGTCTTG	1150
GTCATGAGCT	TGTGGATGA	<u>TGTGAGTTCA</u>	<u>TGAATAGCAT</u>	<u>TCAAACAGTC</u>	1200
<u>AACAGAATAA</u>	<u>CAGCAGCTGA</u>	<u>CTGAGACACA</u>	<u>ATAGTACTAC</u>	GCCAACATGC	1250
TGTGGCTGGA	CTCCACCTAC	CCGACGAACG	AGACCTCTTC	CACCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAGCTCC	GGTGTCCCTG	CTCAGCTTGA	1350
GTCCCAGTCT	TCCAACGCCA	AGGTCGTATA	CTCCAACATC	AAGTTCGGCC	1400
CTATCGGCAG	CACCGGCAAC	TCCAGCGGGG	GTAGCCCTCC	CGGCAGGAGA	1450
AACCCCTCCCG	GTACCACGAC	CACCCGCCGC	CCAGCTACCT	CCACTGGAAAG	1500
CTCTCCCGGC	CCTACTCAGA	CGCACTATGG	CCAGTGCAGT	GGTATTGGGT	1550
ACTCGGGCCC	CACGGTCTGC	GCGAGTGGCA	GCACCTGCCA	GG	1592S

Figure 9: *T. pseudokonigii* genomic DNA sequence.

Figure 10A: *T. pseudokoningii* signal sequence.

QSACTLQTET	HPPLTWQKCS	SGGTCTQQTG	SVVIDANWRW	THATNSSTNC	50
YDGNTWSSTL	CPDNETCAKN	CCLDGAAAYAS	TYGVVTSADS	LSIGFVTQSA	100
QKNVGARLYL	MASDTTYQEF	TLLGNEFSFD	VDVSQLPCGL	NGALYFVSMD	150
ADGGVSKYPT	NTAGAKYGTG	YCDSQLCPRL	KFINGEANVE	GWEPEFSNNAN	200
TGIGGGHGSCC	SEMDIWEANS	ISEALTPHPC	TTVGQEICDG	DSCGGTYSGD	250
RYGGTCDPDG	CDWNPYRLGN	TSFYGPGSSF	ALDTTKKLT	VTQFETSGAI	300
NRYYVQNGVT	FQQPNELGS	YSGNSLDDDY	CAAEEAEFGG	SSFSDKGGLT	350
QFKKATSGGM	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GAVRGSCSTS	400
SGVPAQLESQ	SSNAKVVYSN	IKFGPIGSTG	NSSGGSPPGG	GNPPGTTTR	450
RPATSTGSSP	GPTQTHYGQC	GGIGYSGPTV	CASGSTCQVL	NPYYSQCL	498

Figure 10B: *T. pseudokoningii* mature amino acid sequence.

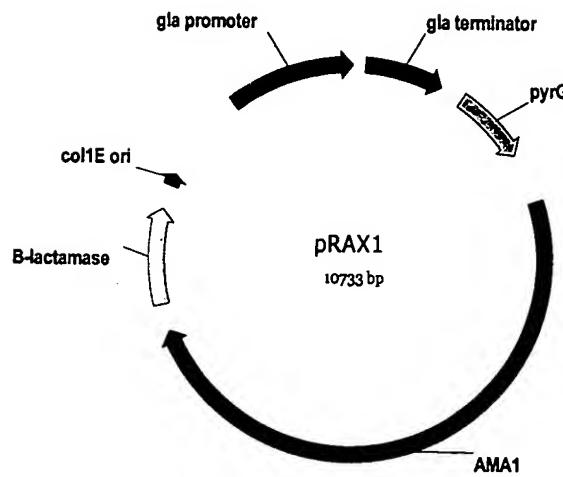


Figure 11: pRAX1

Figure 12: Destination vector pRAXdes2 for expression in *A. niger*

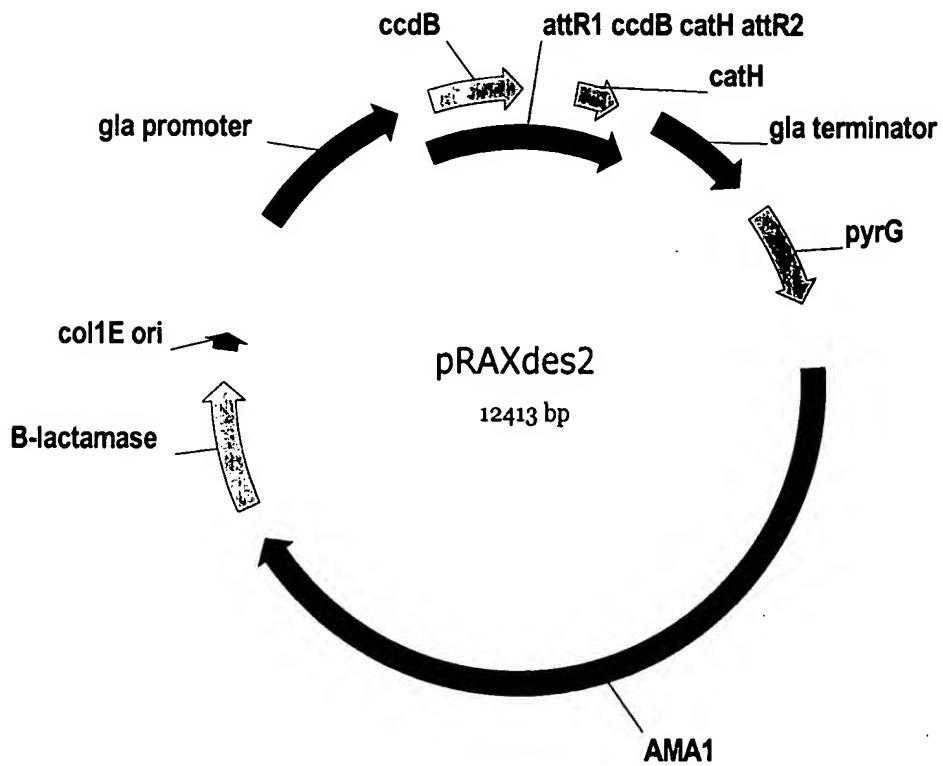


Figure 13: Replicative expression pRAXdesCBH1 vector of CBH1 genes under the control of the glucoamylase promotor.

